

# 运行存储分配

2020年7月4日 21:57

## 1. 运行存储分配

### a. 运行存储分配策略

- i. 编译器在工作过程中，必须为源程序中出现的一些数据对象分配运行时的存储空间
- ii. 对于那些在编译时刻就可以确定大小的数据对象，可以在编译时刻就为它们分配存储空间，这样的分配策略称为静态存储分配（编译时）
- iii. 反之，如果不能在编译时完全确定数据对象的大小，就要采用动态存储分配的策略。即在编译时仅产生各种必要的信息，而在运行时刻，再动态地分配数据对象的存储空间（运行时）

1) 栈式存储分配

2) 堆式存储分配

### b. 运行时动态内存划分

- i. 静态代码区
- ii. 静态数据区
- iii. 动态数据区域
  - 1) 栈(Stack)区
  - 2) 空闲内存
  - 3) 堆(Heap)区（长生命周期数据）

### c. 活动记录

- i. 使用过程(或函数、方法)作为用户自定义动作的单元的语言，其编译器通常以过程为单位分配存储空间
- ii. 过程体的每次执行称为该过程的一个活动(activation)
- iii. 过程每执行一次，就为它分配一块连续存储区，用来管理过程一次执行所需的信息，这块连续存储区称为活动记录(activation record)

### d. 活动记录的一般形式

实参：调用过程传递给被调过程的值

返回值：被调过程返回给调用过程的值

控制链：指向调用者的活动记录（的指针）

- ☒ 访问链：用来访问存放于其它活动记录中的非局部数据  
保存的机器状态

局部数据：被调过程里声明的数据

临时变量：执行过程中产生的临时变量

## 2. 静态存储分配

### a. 在静态存储分配中，编译器为每个过程确定其活动记录在目标程序中的位置

- i. 过程中每个名字的存储位置就确定了
- ii. 因此，这些名字的存储地址可以被编译到目标代码中

- iii. 过程每次执行时，它的名字都绑定到同样的存储单元
- b. 限制条件
  - i. 数组上下界必须是常数
  - ii. 不允许过程的递归调用
  - iii. 不允许动态建立数据实体
  - iv. 如BASIC和FORTRAN都满足这种条件
- c. 顺序分配法
  - i. 按照过程出现的先后顺序逐段分配存储空间
  - ii. 各过程的活动记录占用互不相交的存储空间
  - iii. 处理上简单，但对内存空间的使用不够经济合理
- d. 层次分配法：通过对过程间的调用关系进行分析，凡属无相互调用关系的并列过程，尽量使其局部数据共享存储空间
  - i.  $B[n][n]$ ：过程调用关系矩阵
    - 1)  $B[i][j]=1$ ：表示第*i*个过程调用第*j*个过程
    - 2)  $B[i][j]=0$ ：表示第*i*个过程不调用第*j*个过程
  - ii.  $Units[n]$ ：过程所需内存量矩阵
  - iii.  $base[i]$ ：第*i*个过程局部数据区的基地址
  - iv.  $allocated[i]$ ：第*i*个过程局部数据区是否分配的标志

```
void MakeFortranBlockAllocated(int *Units, int *base, int NoOfBlocks ) {
    /*NoOfBlocks indicating how many blocks there are*/
    int i, j, k, Sum;
    int *allocated; /*used to indicate if the block is allocated*/
    allocated=(int*)malloc(sizeof(int) *NoOfBlocks);
    for(i=0; i<NoOfBlocks; i++) {
        /*Initial arrays base and allocated*/
        base[i]=0;
        allocated[i]=0; }
    for(j=0; j<NoOfBlocks; j++)
        for(i=0; i<NoOfBlocks; i++) {
            Sum=0;
            for(k=0; k<NoOfBlocks; k++)
                Sum+=B[i][k]; /*to check out if block i calls some block which has not been allocated*/
            if(!Sum && !allocated[i]) {
                /*Sum=0 means block i calls no block which has not been allocated;
                allocated [i]=0;means block i is not allocated */
                allocated [i]=1;
                printf("%d:%d -%d \n",i,base[i],base[i]+Units[i]-1);
                for(k=0; k<NoOfBlocks; k++)
                    if(B[k][i]) { /*b[k][i]!=0 means block k calls block i*/
                        /*Since block k calls i, it must be allocated after block i.
                        It means the base of block k must be greater than base of block i*/
                        if(base[k]<base[i]+Units[i])
                            base[k]=base[i]+Units[i];
                        B[k][i]=0; /*Since block i has been allocated B[k][i] should be modified*/ } } }
            free(allocated); }
```

### 3. 栈式存储分配：以栈的形式管理过程运行时刻存储

- a. 当一个过程被调用时，该过程的活动记录被压入栈；当过程结束时，该活动记录被弹出栈
- b. 不仅允许活跃时段不交叠的多个过程调用之间共享空间，而且允许为一个过程编译代码时，它的非局部变量的相对地址总是固定的，和过程调用序列无关
- c. 活动树
  - i. 用来描述程序运行期间控制进入和离开各个活动的情况的树
  - ii. 树中的每个结点对应于一个活动
  - iii. 根结点是启动程序执行的main过程的活动

- iv. 在表示过程p的某个活动的结点上，其子结点对应于被p的这次活动调用的各个过程的活动。按照这些活动被调用的顺序，自左向右地显示它们
  - v. 一个子结点必须在其右兄弟结点的活动开始之前结束
  - d. 活动树与控制栈
    - i. 每个活跃的活动都有一个位于控制栈中的活动记录
    - ii. 活动树的根的活动记录位于栈底
    - iii. 程序控制所在的活动的记录位于栈顶
    - iv. 栈中全部活动记录的序列对应于在活动树中到达当前控制所在的活动结点的路径
  - e. 活动记录的设计原则
    - i. 在调用者和被调用者之间传递的值一般被放在被调用者的活动记录的开始位置，这样它们可以尽可能地靠近调用者的活动记录
    - ii. 固定长度的项（如控制链、访问链、机器状态字）被放置在中间位置
    - iii. 在早期不知道大小的项（如局部、临时数据）被放置在活动记录的尾部
    - iv. 栈顶指针寄存器top\_sp指向活动记录中局部数据开始的位置，以该位置作为基地址
4. 调用序列和返回序列
- a. 过程调用和过程返回都需要执行一些代码来管理活动记录栈，保存或恢复机器状态等
    - i. 调用序列：实现过程调用的代码段。为一个活动记录在栈中分配空间，并在此记录的字段中填写信息
    - ii. 返回序列：恢复机器状态，使得调用过程能够在调用结束之后继续执行
    - iii. 一个调用代码序列中的代码通常被分割到调用过程（调用者）和被调用过程（被调用者）中。返回序列也是如此
  - b. 调用序列
    - i. 调用者计算实际参数的值
    - ii. 调用者将返回地址（程序计数器的值）放到被调用者的机器状态字段中。将原来的top-sp值放到被调用者的控制链中。然后，增加top-sp的值，使其指向被调用者局部数据开始的位置
    - iii. 被调用者保存寄存器值和其它状态信息
    - iv. 被调用者初始化其局部数据并开始执行
  - c. 返回序列
    - i. 被调用者将返回值放到与参数相邻的位置
    - ii. 使用机器状态字段中的信息，被调用者恢复top-sp和其它寄存器，然后跳转到由调用者放在机器状态字段中的返回地址
    - iii. 尽管top-sp已经被减小，但调用者仍然知道返回值相对于当前top-sp值的位置。因此，调用者可以使用那个返回值
  - d. 调用与被调用者间的任务划分
    - i. 调用者填写：参数、控制链、机器状态
    - ii. 被调用者填写：返回值、机器状态

- iii. 调用者更新，被调用者回复：top\_sp
- e. 变长数据的存储分配
  - i. 现代程序设计语言中，在编译时刻不能确定大小的对象将被分配在堆。但如果它们是过程的局部对象，也可以将它们分配在运行时刻栈中
  - ii. 将对象放置在栈区的原因：可以避免对它们的空间进行垃圾回收，也就减少了相应的开销
  - iii. 只有一个数据对象局部于某个过程，且当此过程结束时它变得不可访问，才可以使用栈为这个对象分配空间
  - iv. 如局部数组常存在栈区。但不写进活动记录，活动记录中只有其指针
- 5. 非局部数据的访问
  - a. 一个过程除了可以使用过程自身定义的局部数据以外，还可以使用过程外定义的非局部数据
  - b. 不支持过程嵌套声明的语言：不可以在一个过程中声明另一个过程。例：C
    - i. 过程中使用的数据要么是自身定义的局部数据，要么是在所有过程之外定义的全局数据
    - ii. 全局变量被分配在静态区，使用静态确定的地址访问它们
    - iii. 其它变量一定是栈顶活动的局部变量。可以通过运行时刻栈的top\_sp指针访问它们
  - c. 支持过程嵌套声明的语言：可以在一个过程中声明另一个过程。例：Pascal
    - i. 一个过程除自身定义的局部数据和全局定义的数据以外，还可以使用外围过程中声明的对象
    - ii. 过程的嵌套深度：不内嵌在任何其它过程中的过程，设其嵌套深度为1；如果过程p在嵌套深度为i的过程中定义，则设定p的嵌套深度为i+1
    - iii. 变量的嵌套深度：变量声明所在过程的嵌套深度
  - d. 访问链(Access Links)
    - i. 静态作用域规则：只要过程b的声明嵌套在过程a的声明中，过程b就可以访问过程a中声明的对象
    - ii. 可以在相互嵌套的过程的活动记录之间建立一种称为访问链的指针，使得内嵌的过程可以访问外层过程中声明的对象
    - iii. 如果过程b在源代码中直接嵌套在过程a中(即b的嵌套深度比a的嵌套深度多1时)，那么b的任何活动中的访问链都指向最近的a的活动
  - e. 访问链的建立
    - i. 建立访问链的代码属于调用序列的一部分
    - ii. 假设嵌套深度为nx的过程x调用嵌套深度为ny的过程y( $x \rightarrow y$ )
    - iii.  $n_x < n_y$ 的情况(外层调用内层)：y一定是直接在x中定义的(例如： $s \rightarrow q$ ,  $q \rightarrow p$ )，因此， $n_y = n_x + 1$ 。在调用代码序列中增加一个步骤：在y的访问链中放置一个指向x的活动记录的指针
    - iv.  $n_x = n_y$ 的情况(本层调用本层)：递归调用(例如： $q \rightarrow q$ )。被调用者的活动记录的访问链与调用者的活动记录的访问链是相同的，可以直接复制
    - v.  $n_x > n_y$ 的情况(内层调用外层，如： $p \rightarrow e$ )：过程x必定嵌套在某个过程z中，

而z中直接定义了过程y。从x的活动记录开始，沿着访问链经过 $n_x - n_y + 1$ 步就可以找到离栈顶最近的z的活动记录。y的访问链必须指向z的这个活动记录

## 6. 堆式存储分配

- a. 堆式存储分配是把连续存储区域分成块，当活动记录或其它对象需要时就分配
  - i. 块的释放可以按任意次序进行，所以经过一段时间后，可能包含交错的正在使用和已经释放的区域
- b. 申请：设当前自由块总长为M，欲申请长度为n
  - i. 如果存在若干个长度大于或等于n的存储块，可按以下策略之一分配
    - 1) 取长度m满足需求的最小的自由块
    - 2) 取长度m满足需求的最大的自由块
    - 3) 取长度m满足需求的第1个自由块，将长度为m-n的剩余部分仍放在自由链中
  - ii. 如果不存在长度大于或等于n的存储块
    - 1) 如果 $M \geq n$ ，将自由块在堆中进行移位和重组（对各有关部分都需作相应的修改，是一件十分复杂和困难的工作）
    - 2) 如果 $M < n$ ，则应采用更复杂的策略来解决堆的管理问题
- c. 释放：只需将被释放的存储块作为新的自由块插入自由链中，并删除已占块记录表中相应的记录即可
- d. 为实现堆式存储管理，须完成大量的辅助操作。如排序、查表、填表、插入、删除、.....其空间和时间的开销较大

## 7. 符号表

- a. 符号表的组织：为每个作用域（程序块）建立一个独立的符号表
  - i. 为每个过程或作用域建立的符号表与编译时的活动记录是对应的。一个过程的非局部名字的信息可以通过扫描外围过程的符号表而得到
- b. 标识符的基本处理方法
  - i. 当在某一层的声明语句中识别出一个标识符(id的定义性出现)时，以此标识符查相应于本层的符号表
    - 1) 如果查到，则报错并发出诊断信息“id重复声明”
    - 2) 否则，在符号表中加入新登记项，将标识符及有关信息填入
  - ii. 当在可执行语句部分扫视到标识符时(id的应用性出现)
    - 1) 首先在该层符号表中查找该id，如果找不到，则到直接外层符号表中去查.....，一旦找到，则在表中取出有关信息并作相应处理
    - 2) 如果查遍所有外层符号表均未找到该id，则报错并发出诊断信息“id未声明”

## 8. 符号表的建立

- a. 在允许嵌套过程声明的语言中，局部于每个过程的名字可以使用上一章介绍的方法分配相对地址；当看到嵌套的过程p时，应暂时挂起对外围过程q声明语句的处理
- b. 嵌套过程声明语句的SDT
  - i. 需要配合符号表指针栈tblptr和符号表内相对偏移位置栈offset使用
  - ii. mktable(previous) 创建一个新的符号表，并返回指向新表的指针，参数

previous指向先前创建的符号表(外围过程的符号表)。常用于在产生式中新过程前的 $M \rightarrow \epsilon$ 里

- iii. addwidth(table, width) 将table指向的符号表中所有表项的宽度之和width记录在符号表的表头中
- iv. enterproc(table, name, newtable) 在table指向的符号表中为过程name建立一条记录, newtable指向过程name的符号表
- v. enter(table, name, type, offset) 在table指向的符号表中为名字name建立一个新表项, 使type和offset的值增加